**Protein Engineering** vol.14 no.11 pp.881–890, 2001

# Homology modelling of the human eukaryotic initiation factor 5A (eIF-5A)

**Angelo M.Facchiano[1,2,3,4], Paola Stiuso[1,2], Maria Luisa Chiusano[1,2], Michele Caraglia[1], Gaia Giuberti[1], Monica Marra[1], Alberto Abbruzzese[1] and Giovanni Colonna[1,2]**

[1]Dipartimento di Biochimica e Biofisica, [2]Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università di Napoli, via Costantinopoli 16, 80138 Napoli and [3]Istituto di Scienze dell'Alimentazione, CNR, via Roma 52A/C, 83100 Avellino, Italy

[4]To whom correspondence should be addressed.

**Homology modelling of the human eIF-5A protein has been performed by using a multiple predictions strategy. As the sequence identity between the target and the template proteins is nearly 30%, which is lower than the commonly used threshold to apply with confidence the homology modelling method, we developed a specific predictive scheme by combining different sequence analyses and predictions, as well as model validation by comparison to structural experimental information. The target sequence has been used to find homologues within sequence databases and a multiple alignment has been created. Secondary structure for each single protein has been predicted and compared on the basis of the multiple sequence alignment, in order to evaluate and adjust carefully any gap. Therefore, comparative modelling has been applied to create the model of the protein on the basis of the optimized sequence alignment. The quality of the model has been checked by computational methods and the structural features have been compared to experimental information, giving us a good validation of the reliability of the model and its correspondence to the protein structure in solution. Last, the model was deposited in the Protein Data Bank to be accessible for studies on the structure–function relationships of the human eIF-5A.**

*Keywords*: consensus of methods/eIF-5A/homology modelling/
model validation/prediction strategy

## Introduction

The knowledge of detailed structural organization is crucial in understanding the role of proteins in the cell and the related molecular mechanisms. Genome sequencing projects continuously detect new protein sequences, thus providing new information for the application of computational methods, stimulating the improvement of this field, which represents a good alternative to the relatively slow experimental processes for predicting protein structure (Rodriguez *et al*., 1998; Westhead and Thornton, 1998). The comparative assessment of techniques of protein structure prediction, CASP (see http://PredictionCenter.llnl.gov), is a bi-annual state-of-the-art check for the protein structure prediction community. The three prediction strategies currently evaluated are comparative modelling (or homology modelling), fold recognition and *ab initio* methods. Comparative modelling, which is considered the most reliable method to predict the three-dimensional (3D) structure of a protein, creates the model of a 'target' protein on the 'template' coordinates of homologous proteins, on the basis that the same structure is conserved within a protein family (Flores *et al*., 1993; Sali and Blundell, 1993). It must be remarked that this method can give very good results, but it can be applied only under specific conditions. There is wide agreement on the observation that comparative modelling can produce wrong models if low sequence identity exists between the target protein and the chosen template structure; as well, the risk of errors must be taken into account when the sequence similarity falls in the medium range. The lowest identity to be used in homology modelling is assessed at 40% (Rodriguez *et al*., 1998) or at 20% (Westhead and Thornton, 1998). Evaluations of CASP2 and CASP3 results suggest that modelling based on homologues with 30% identity or less produce not useful and low quality models (Martin *et al*., 1997; Sternberg *et al*., 1999). Therefore, comparative modelling starting from sequence identities within the 20–40% range may be wrong as a consequence of the low quality of the sequence alignment between the target and the template protein. Low identity can mean different alignments with very similar scores and the best of them are not necessarily suitable for the comparative modelling procedure, as a consequence of insertion of gaps within secondary structure elements or the need for too many or too long gaps. Nevertheless, when functional information allows the definition of two proteins as 'homologues', although they have low sequence similarity, it is still reasonable to consider that the same function is played by the same structure and to try to apply the comparative modelling strategy. The critical point is the assessment of a right and suitable sequence alignment. It is a common procedure to improve by visual inspection the sequence alignment obtained with computer software in order to consider information such as the position of secondary structure elements or the involvement of specific amino acids in the protein function. In this paper we propose a predictive scheme to be applied in such cases, consisting of: (i) the assessment of the suitable sequence alignment by comparison and consensus among pair sequence alignments, multiple sequence alignments and secondary structure predictions by independent methods; (ii) the comparative modelling on the basis of the chosen alignment, inclusive of checks for the stereo-chemical quality of the predicted model; (iii) the comparison of secondary structure features of the model with secondary structure predictions; (iv) the comparison of structural features of the model with experimental information about the structure and the function of the protein. At least two different advantages are included in this scheme. The first is the search for the consensus of different approaches, for both the sequence alignments and the structure predictions. This is considered a good way to improve the reliability of results obtained by computational analyses based on the simple protein sequence. The second one is the validation of the model by comparison to experimental information about the structure of the protein, like secondary structure content by circular dichroism (CD), surface exposure of specific amino acids and others obtained by investigation with spectrophotometry and fluorescence methods. Such experimental results represent structural details which can not easily be used in prediction methods, but can be exploited to validate the predicted model.

We applied our prediction strategy to the human eukaryotic translation initiation factor 5A (eIF-5A), a 18-kDa protein, highly conserved from yeast to mammalian cells (Park *et al.*, 1984; Gordon *et al.*, 1987). eIF-5A precursor [(ec-eIF-5A(lys)] is the only cellular protein known to contain a specific lysine residue which is transformed into the unique amino acid hypusine [$N^{\varepsilon}$-(4-amino-2-hydroxybutyl)-lysine] by a series of post-translational reactions: (i) the transfer of the butylamine moiety from spermidine to the ε-amino group of one of the lysine residues in the eIF-5A precursor protein, thus forming peptide-bound deoxyhypusine (Park *et al.*, 1982); (ii) the intermediate hydroxylation at C-2 of the incoming 4-aminobutyl moiety to form hypusine (Abbruzzese *et al.*, 1985, 1986, 1988a,b). eIF-5A promotes the formation of the first peptide bond during the initial stage of protein synthesis (Hersey, 1991). However, the actual *in vivo* function of eIF-5A is to date still only partially known. eIF-5A precursors which do not contain hypusine, have little, if any, activity (Park *et al.*, 1991). In addition, the lysine→arginine variant is unable to stimulate methionyl-puromycin synthesis *in vitro* (Hersey *et al.*, 1990; Park *et al.*, 1991) and is inactive *in vivo* (Smit-McBride *et al.*, 1989; Beninati *et al.*, 1995). Therefore, hypusine synthesis is required for the biological activity of the protein. Moreover, the ec-eIF-5A(lys) modification is correlated with cell proliferation (Abbruzzese, 1988; Abbruzzese *et al.*, 1988a,b; Beninati *et al.*, 1990; Schnier *et al.*, 1991; Caraglia *et al.*, 1997), and agents that block the lysine–hypusine transformation (Park *et al.*, 1984; Abbruzzese *et al.*, 1989, 1991; Jakus *et al.*, 1993) inhibit the growth of mammalian cells (Park *et al.*, 1993) inducing reversible arrest at the $G_1$-S boundary of the cell cycle (Park *et al.*, 1981; Cooper *et al.*, 1982; Abbruzzese *et al.*, 1986, 1988; Park, 1987; Lalande and Hanausske-Abel, 1990; Beninati *et al.*, 1990, 1993). The polyamine-dependent modification of eIF-5A has been related to the triggering of apoptosis in tumour cells (Tome and Gerner, 1997; Tome *et al.*, 1997). It has also been reported that eIF-5A can accumulate at nuclear pore-associated intranuclear filaments in mammalian cells (Rosorius *et al.*, 1999) and interacts with the general nuclear export receptor CRM1 being transported from the nucleus to the cytoplasm (Rosorius *et al.*, 1999). These findings open a new scenario in which eIF-5A may also function as a nucleocytoplasmic shuttle protein of mRNAs eventually correlated with cell proliferation and apoptosis.

The knowledge of the 3D structure of human eIF-5A can help in the understanding of its function and role in the cell, in order to study the molecular interaction with other proteins, as well as to design new molecules useful to modulate its activity. Different researchers investigated the structure of this human protein, both by experimental methods (Joao *et al.*, 1995; Klier *et al.*, 1995; Stiuso *et al.*, 1999) and predictive approaches (Gerloff *et al.*, 1998). However, these works are still far from giving us a 3D model of the protein. Recently, the 3D structures of two eIF-5A from archea have been solved (Kim *et al.*, 1998; Peat *et al.*, 1998). Their sequence similarity to human eIF-5A is lower than 40%, so that the comparative modelling strategy must be applied with caution, it being possible to obtain a wrong model as a consequence of a wrong sequence alignment. For this reason we chose this protein to apply our predictive scheme. This work provides an example of combining different predictive methods and experimental results, in order to develop a new prediction strategy, inclusive of an evaluation of its reliability. Finally, the 3D model of the human eIF-5A represents an essential support to further studies on its structure–function relationships.

## Methods

### *In silico: sequence and 3D structure modelling and analysis*

The sequence of human eIF-5A has been taken from the SWISSPROT databases (ID code IF5A_HUMAN). Structure prediction of human eIF-5A has been based on the availability of the 3D models of the homologous protein from *Methanococcus jannaschii* (Kim *et al.*, 1998), PDB code 1EIF and from *Pyrobaculum aerophilum* (Peat *et al.*, 1998), PDB code 1BKB. The search for sequence similarity within databases has been performed with the BLAST program (Altschul *et al.*, 1997). The alignment of eIF-5A sequences has been performed by the PILEUP program of the GCG package and by CLUSTALW (Thompson *et al.*, 1994). Secondary structure was predicted with the SOPMA method (Geourjon and Deleage, 1994, 1995) which considers the consensus of different prediction methods and does not perform multiple sequence alignment. This feature is important in our approach because we need to compare independent alignments of sequences and secondary structures. PHD (Rost and Sander, 1993) and JPRED (Cuff *et al.*, 1998) were also used for further comparison. The secondary structure of 3D models has been assigned with the program DSSP (Kabsch and Sander, 1983). The programs MODELLER (Sali and Blundell, 1992) and Quanta (Molecular Simulations, Inc., San Diego, CA) were used to build 10 full-atom models of human eIF-5A according to the comparative protein modelling method. The stereochemical quality of the models was verified with the program PROCHECK (Laskowski *et al.*, 1993), in order to select the best model. The search for structural classification of bacterial eIF-5A was performed on SCOP (Murzin *et al.*, 1995) and CATH (Orengo *et al.*, 1997) databases. The solvent accessibility of amino acids was evaluated by the program NACCESS (Hubbard *et al.*, 1991) calculating the atomic accessible surface defined by rolling a probe of 1.40 Å around the van der Waals surface of the protein model. Model figures were drawn with the InsightII package (Molecular Simulations, Inc.)

### *In vitro: structural investigation*

*Protein preparation*. eIF-5A precursor(Lys) was prepared by over-expression of human eIF-5A-cDNA in *Escherichia coli* according to the method of Smit-McBride *et al.* (Smit-McBride *et al.*, 1989). The protein is homogeneous when tested by SDS–PAGE showing a single band.

*Protein concentration*. Evaluation of protein concentration was performed spectrophotometrically by using $\varepsilon_{275} = 4100$ $M^{-1}$ $cm^{-1}$. This value was calculated by the content of tyrosine and phenylalanine residues by using molar extinction coefficients of 1250 and 200 $M^{-1}$ $cm^{-1}$, respectively (Wetlaufer, 1962). This method has been found very reliable for this purpose with a maximum error of 5% (Gill and von Hippel, 1989). The Bradford method was also used (Bradford, 1976).

*Cysteine titration*. Sulfydryl group titration of native and denatured protein was performed by the 5,5′-dithiobis (2-nitrobenzoic acid) (DTNB) method (Ellman, 1959). A large excess of DTNB is added to the protein and the increase of absorbance is monitored spectrophotometrically by using $\varepsilon_{412} = 13\,600$ M$^{-1}$ cm$^{-1}$ for the DTNB–sulfur adduct (Ellman, 1959). To analyse the total content of sulfydryl groups, the protein sample was denatured in 5 M guanidinium chloride (GdnHCl) for 24 h. After the addition of DTNB, the absorbance at 412 nm increased quickly, showing a monophasic transition, up to a value corresponding to four DTNB–sulfur adducts. A similar protocol has been applied to the native protein. The absorbance values have been reported as the number of cysteine residues titrated.

*Spectral measurement*. CD measurements were carried out on a Jobin Yvon Mark III spectropolarimeter equipped with a temperature-controlled cell holder, in 0.05 M Tris, pH 7.8 containing 0.15 M KCl. Mean residue ellipticities were calculated by:

$$[\theta]_\lambda = (\mathrm{MRW}q_{\mathrm{obs}})/(10lc)$$

where $[\theta]_\lambda$ is the mean residue ellipticity (deg cm$^2$/dmol) at a particular wavelength, $q_{\mathrm{obs}}$ is the observed ellipticity, MRW is the mean residue molecular weight calculated from the sequence, $l$ is the optical path length (cm) and $c$ is the concentration (g/ml). Cells between 0.01 and 1 cm were used in the far-UV. CD spectra were measured by computer so as to have reliable tracings even in the case of very diluted protein solutions, in addition to eliminating artefacts and obtaining good blank correction. The CD spectra were analysed in the region between 200 and 250 nm to evaluate the amount of secondary structure. A computer program with different methods of analysis (Menéndez-Arias *et al*., 1988) was used.

The protein samples under spectroscopic investigation were always dialysed against numerous changes of buffer, which was used as a blank for the measurements. It was reported that the eIF-5A precursor has a tendency to dimerize reversibly (Chung *et al*., 1991). According to these authors, the low protein concentration used in our experiments is populated by the monomeric form (dissociated state) of the protein. In fact, no concentration dependence was detected under our experimental conditions.

*Chemicals and solutions*. Spectroscopic grade GdnHCl was from Schwarz-Mann. All chemicals were reagent grade and were purchased from BDH (British Drug Houses, UK). In denaturation experiments the protein was added to buffered solutions of GdnHCl; 0.15 M KCl was present in all solutions. The spectroscopic measurements were then followed until an apparent equilibrium was reached. pH measurements were carried out by using combination electrodes with a Radiometer pH meter.

## Results

The sequence of human eIF-5A has been analysed by computer programs in order to find similar sequences in databases and perform structure predictions. Table I summarizes the results of the BLAST search within the non-redundant database of all proteins sequences. By considering the best alignments found (i.e. an *E* value better than $10^{-6}$), the human eIF-5A sequence is significantly similar to many proteins defined as ‘initiation factor 5A’, thus confirming that this protein is well conserved within different organisms, in both eukaryotic (sequences 1–35) and archea (sequences 36–48) groups. Sequence #44, i.e. a putative 20-kDa subunit of the V-ATPase from *Neurospora crassa* (an eukaryotic organism), is the only exception. Its presence will be examined in more detail below. The search has also evidenced the presence of two experimental 3D models of IF-5A precursor from archea, i.e. from *M.jannaschii* (PDB code: 1EIF and 2EIF) and *P.aerophilum* (PDB code: 1BKB), to be considered for the comparative modelling. The BLAST alignments (Figure 1A) do not include any gaps and show that the most similar region includes the lysine/hypusine residue, as well as other matches observed along the whole sequence. The amino acid identities between human and the other sequences represent 34 and 32% of the alignments to the two archea sequences and the amino acid similarities represent 54 and 56%. Multiple alignment of the three sequences has been performed with the PILEUP program (Figure 1B). Identical residues within the three aligned sequences represent approximately 16% of the human sequence, whereas the sum of identical and similar residues represent 40% of the human sequence. The lower identity and similarity evidenced by PILEUP as compared to the BLAST alignments, as well as the addition of few gaps, are evidently due to multiple alignment features.

As the sequence identity of human eIF-5A to both possible template proteins was lower than the threshold value of 40%, the sequence alignment was assessed with care in order to apply the comparative modelling strategy. We performed multiple alignment and also secondary structure predictions in order to verify that the eIF-5A protein family is well conserved, so that similar structures can be assumed for the human and the two archea proteins, as well as to improve the alignment by consideration of the possible position of secondary structure elements. We aligned the 48 sequences listed in Table I with the program CLUSTALW (Figure 2). An identity matrix, showing the proportion of identical residues between all of the sequences in the alignment reported in Figure 2, has been created (data not shown, see supplemental material). The average of the percentage of identities is 42%. Moreover, the consensus sequence has been created by choosing the most frequent amino acid for each column in the alignment, leaving an empty space where no consensus has been found. The consensus sequence (bottom row in Figure 2) shows that approximately 45% of amino acids is conserved (69/153). Both the multiple alignment and the consensus show a good similarity level among the sequences considered, with the most conserved region in the proximity of the lysine/hypusine residue. Also evident is the high similarity in the two distinct groups, i.e. sequences from the eukaryotes (1–35) and the archea (36–48). Sequence #44 does not seem to be a member of the IF5A family, although it was found by BLAST to have an *E* value better than some archea IF5A sequences; therefore, it has been included in this alignment to evaluate its similarity. It appears not to be well integrated into the multiple alignment and its presence inserts some more gaps in all the other sequences. When sequence #44 was removed and the multiple alignment performed again (not shown), these gaps were removed and no other difference appears between the two alignments.

The secondary structure has been predicted for each of the 48 proteins and aligned according to the primary structure multiple alignment (Figure 3). High consensus results in the global alignment, mainly in the proximity of the lysine/hypusine residue and the middle of the sequences but also

**Table I.** Summary of BLAST results

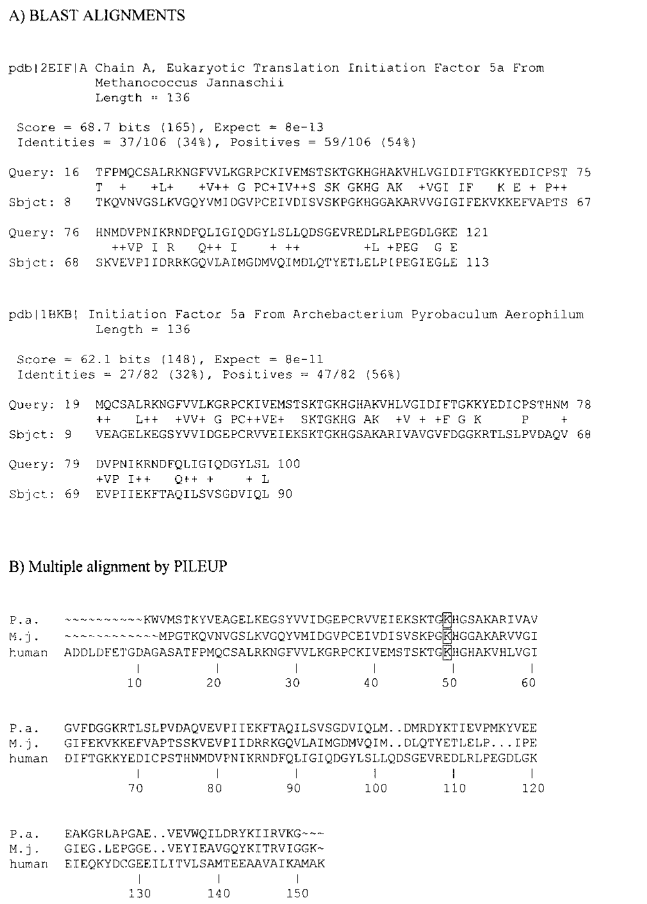
| # | DB[a] | ACC | Group[b] | Organism[c] | *E* value |
|---|---|---|---|---|---|
| 1 | Gp | NP_001961.1 | Euk | *Homo sapiens* | 8.00E–85 |
| 2 | Sp | P10160 | Euk | *Oryctolagus cuniculus* | 1.00E–83 |
| 3 | Pir | I53801 | Euk | *Homo sapiens* | 4.00E–82 |
| 4 | Sp | Q07460 | Euk | *Gallus gallus* | 8.00E–72 |
| 5 | Gb | AAF13315.1 | Euk | *Spodoptera exigua* | 2.00E–61 |
| 6 | Gb | AAF47151.1 | Euk | *Drosophila melanogaster* | 9.00E–57 |
| 7 | Sp | P56289 | Euk | *Schizosaccharomyces pombe* | 3.00E–56 |
| 8 | Pir | T40248 | Euk | *Schizosaccharomyces pombe* | 7.00E–56 |
| 9 | Sp | O94083 | Euk | *Candida albicans* | 1.00E–55 |
| 10 | Gp | NP_010880.1 | Euk | *Saccharomyces cerevisiae* | 2.00E–55 |
| 11 | Gp | NP_012581.1 | Euk | *Saccharomyces cerevisiae* | 1.00E–54 |
| 12 | Sp | P13651 | Euk | *Dictyostelium discoideum* | 2.00E–52 |
| 13 | Sp | Q09121 | Euk | *Gallus gallus* | 8.00E–51 |
| 14 | Sp | P38672 | Euk | *Neurospora crassa* | 2.00E–49 |
| 15 | Sp | Q20751 | Euk | *Caenorhabditis elegans* | 6.00E–49 |
| 16 | Pir | S41010 | Euk | *Caenorhabditis elegans* | 2.00E–46 |
| 17 | Sp | P34563 | Euk | *Caenorhabditis elegans* | 2.00E–46 |
| 18 | Pir | B42156 | Euk | *Gallus gallus* | 1.00E–45 |
| 19 | Gb | AAD39281.1 | Euk | *Arabidopsis thaliana* | 1.00E–44 |
| 20 | Sp | P56335 | Euk | *Solanum tuberosum* | 4.00E–43 |
| 21 | Sp | P56336 | Euk | *Solanum tuberosum* | 8.00E–43 |
| 22 | Sp | P24922 | Euk | *Nicotiana plumbaginifolia* | 1.00E–42 |
| 23 | Gb | AAF27938.1 | Euk | *Euphorbia esula* | 3.00E–42 |
| 24 | Gb | AAC67555.1 | Euk | *Oryza sativa* | 5.00E–42 |
| 25 | Sp | P56333 | Euk | *Solanum tuberosum* | 5.00E–42 |
| 26 | Sp | P56332 | Euk | *Zea mays* | 9.00E–42 |
| 27 | Sp | P24921 | Euk | *Nicotiana tabacum* | 8.00E–41 |
| 28 | Sp | P56337 | Euk | *Solanum tuberosum* | 1.00E–40 |
| 29 | Pir | S21058 | Euk | *Nicotiana tabacum* | 1.00E–39 |
| 30 | Emb | CAB65463.1 | Euk | *Senecio vernalis* | 2.00E–39 |
| 31 | Sp | P26564 | Euk | *Medicago sativa* | 3.00E–39 |
| 32 | Gb | AAB21933.1 | Euk | *Gallus gallus* | 3.00E–37 |
| 33 | Gb | AAB21928.1 | Euk | *Gallus gallus* | 4.00E–35 |
| 34 | Sp | Q26571 | Euk | *Schistosoma mansoni* | 4.00E–17 |
| 35 | Pir | C64453 | Arche | *Methanococcus jannaschii* | 3.00E–11 |
| 36 | Sp | Q58625 | Arche | *Methanococcus jannaschii* | 3.00E–11 |
| 37 | Pdb | 2EIF | Arche | *Methanococcus jannaschii* | 3.00E–11 |
| 38 | Pdb | 1EIF | Arche | *Methanococcus jannaschii* | 3.00E–11 |
| 39 | Gb | AAB21930.1 | Arche | *Aeropyrum pernix* | 5.00E–11 |
| 40 | Sp | Q9YA53 | Arche | *Aeropyrum pernix* | 6.00E–11 |
| 41 | Sp | P28461 | Arche | *Sulfolobus acidocaldarius* | 1.00E–10 |
| 42 | Sp | O29612 | Arche | *Archaeoglobus fulgidus* | 2.00E–10 |
| 43 | Sp | P56635 | Arche | *Pyrobaculum aerophilum* | 2.00E–09 |
| 44 | Gb | AAB61278.1 | Euk | *Neurospora crassa*[c] | 2.00E–09 |
| 45 | Pdb | 1BKB | Arche | *Pyrobaculum aerophilum* | 3.00E–09 |
| 46 | Pir | H75120 | Arche | *Pyrococcus abyssi* | 5.00E–08 |
| 47 | Sp | O50089 | Arche | *Pyrococcus horikoshii* | 5.00E–08 |
| 48 | Sp | O26955 | Arche | *Methanobacterium thermoautotrophicum* | 8.00E–07 |

[a] The sequence data banks are reported according to the BLAST code: Emb, EMBL; Gb, Genebank; Gp, Genpept; Sp, Swissprot.

[b] Group refers to eukaryotic (euk) or archea organisms.

[c] All sequences found are initiation factor 5A, with the only exception of sequence #44, which is a putative 20-kDa subunit of the V-ATPase.

at the extremities, and the presence of beta structure is predominant for each protein. Similar secondary structure topology is also predicted for sequence #44. Few helices are predicted for many proteins at both extremities. It should be observed that the few gaps in the alignment fall in loop regions or near the extremities of secondary structure elements, but never in the middle of strands or helices. The same consideration concerns the experimental secondary structure, assigned to the PDB structures with the DSSP program and also reported in Figure 3 (bottom). Two gaps result at the extremities of two different strands, but they were not influent on the comparative modelling because they will be removed in the next step (see below). The alignment of the predicted secondary structures gives us a confirmation of the good quality of the multiple alignment of the protein sequences (see also in Discussion). The global similarity highlighted by the alignments, as well as the similar function, gave us the confidence to apply the comparative protein modelling method to build the 3D model of human eIF-5A on the two homologue template structures from archea. The two template structures, as reported by the FSSP database (http://www2.ebi.ac.uk/dali/fssp/), have 43% sequence identity and a high 3D structure similarity (RMSD = 1.5). As shown in Figure 3 (bottom) the secondary structure assigned by DSSP is very similar (see also Table II), the main difference being a short helix (four residues) in the C-terminal domain of *P.aerophilum* IF5A, not defined as a helix by DSSP in *M.jannaschii* IF5A. The superimposition of the two PDB structures does not show significant differences (data not shown, see supplemental material). We used the alignment of these three sequences, taken from the alignment



A) BLAST ALIGNMENTS

```
pdb|2EIF|A Chain A, Eukaryotic Translation Initiation Factor 5a From
           Methanococcus Jannaschii
           Length = 136

 Score = 68.7 bits (165), Expect = 8e-13
 Identities = 37/106 (34%), Positives = 59/106 (54%)

Query: 16  TFPMQCSALRKNGFVVLKGRPCKIVEMSTSKTGKHGHAKVHLVGIDIFTGKKYEDICPST 75
            T  +   +L+   +V++ G PC+IV++S SK GKHG AK  +VGI IF   K E + P++
Sbjct: 8   TKQVNVGSLKVGQYVMIDGVPCEIVDISVSKPGKHGGAKARVVGIGIFEKVKKEFVAPTS 67

Query: 76  HNMDVPNIKRNDFQLIGIQDGYLSLLQDSGEVREDLRLPEGDLGKE 121
              ++VP I R   Q++ I    + ++        +L +PEG  G E
Sbjct: 68  SKVEVPIIDRRKGQVLAIMGDMVQIMDLQTYETLELPIPEGIEGLE 113


pdb|1BKB| Initiation Factor 5a From Archebacterium Pyrobaculum Aerophilum
           Length = 136

 Score = 62.1 bits (148), Expect = 8e-11
 Identities = 27/82 (32%), Positives = 47/82 (56%)

Query: 19  MQCSALRKNGFVVLKGRPCKIVEMSTSKTGKHGHAKVHLVGIDIFTGKKYEDICPSTHNM 78
            ++   L++  +VV+ G PC++VE+  SKTGKHG AK  +V + +F G K     P    +
Sbjct: 9   VEAGELKEGSYVVIDGEPCRVVEIEKSKTGKHGSAKARIVAVGVFDGGKRTLSLPVDAQV 68

Query: 79  DVPNIKRNDFQLIGIQDGYLSL 100
            +VP I++   Q++ +    + L
Sbjct: 69  EVPIIEKFTAQILSVSGDVIQL 90
```

B) Multiple alignment by PILEUP

```
P.a.   ~~~~~~~~~~KWVMSTKYVEAGELKEGSYVVIDGEPCRVVEIEKSKTGKHGSAKARIVAV
M.j.   ~~~~~~~~~~~~MPGTKQVNVGSLKVGQYVMIDGVPCEIVDISVSKPGKHGGAKARVVGI
human  ADDLDFETGDAGASATFPMQCSALRKNGFVVLKGRPCKIVEMSTSKTGKHGHAKVHLVGI
                |         |         |         |         |         |
               10        20        30        40        50        60

P.a.   GVFDGGKRTLSLPVDAQVEVPIIEKFTAQILSVSGDVIQLM..DMRDYKTIEVPMKYVEE
M.j.   GIFEKVKKEFVAPTSSKVEVPIIDRRKGQVLAIMGDMVQIM..DLQTYETLELP...IPE
human  DIFTGKKYEDICPSTHNMDVPNIKRNDFQLIGIQDGYLSLLQDSGEVREDLRLPEGDLGK
                |         |         |         |         |         |
               70        80        90       100       110       120

P.a.   EAKGRLAPGAE..VEVWQILDRYKIIRVKG~~~
M.j.   GIEG.LEPGGE..VEYIEAVGQYKITRVIGGK~
human  EIEQKYDCGEEILITVLSAMTEEAAVAIKAMAK
                |         |         |
               130       140       150
```

**Fig. 1.** (**A**) Pairwise alignments of human eIF-5A to the *M.jannaschii* and *P.aerophilum* IF-5A sequences, as a result of the BLAST search. (**B**) Multiple alignment of the *M.jannaschii (*M.j.), *P.aerophilum* (P.a.) and human IF-5A sequences. The alignment was created by the PILEUP program (GCG package).

of 48 sequences, with few manual adjustments to eliminate gaps present in all three sequences, thus removing the two breaks at the extremities of strands. Ten structural models of human eIF-5A have been created by using the Modeller program (Sali and Blundell, 1993). The backbone of 10 models overlapped well and the most reliable structure was chosen by evaluating the stereochemical quality of the models using the PROCHECK package (Laskowski *et al.*, 1993). The model has been submitted to the PDB and accepted (PDB code: 1FH4) having passed the required processing for validation.

The secondary structure topology assigned to the predicted model is in good agreement with results of secondary structure prediction (see Figure 3, bottom). Table II summarizes the secondary structure content by predictions and experimental methods. The representation of the protein structure, with schematized secondary structure, is shown in Figure 4. It is well evident that the global structure (Figure 4A) consists of two all-beta domains, in agreement with the structural classifications reported by SCOP (Murzin *et al.*, 1995) and CATH (Orengo *et al.*, 1997) databases for the microbial eIF-5A proteins. According to the CATH database, the N-terminal domain can be defined as a roll architecture (Figure 4B), whereas the C-terminal domain corresponds to the barrel architecture (Figure 4C). The N-terminal segment 1–13 is unordered, as a consequence of the lack of correspondent segments in the template structures. Although secondary structure predictions suggest this segment is a coil, it may be possible to assume a helical structure (see Discussion). It is interesting to note that the N-terminal domain (segment 14–73, by excluding the not modelled N-terminal segment) has a lower content of unordered structure, i.e. 17%, when compared to the 25% observed in the C-terminal domain (segment 74–128). This aspect will be discussed below.

The exposure of Lys49 has been evaluated by visual analysis of the model, as well as by specific programs. This amino acid is well exposed and seems to be in good agreement with the possible requirements for its post-transcriptional modification to hypusine (see Figure 4B).

The model has been compared to experimental results in order to verify the reliability of the predicted structure.

Figure 5 shows the far-UV CD spectra of the native protein. The spectrum shows a well pronounced minimum centred at approximately 205 nm, and an evident shoulder at 215–216 nm. This last spectral feature suggests the presence of β-sheet structure which is characterized by a dichroic band with a minimum at 216 nm. The relatively low intensity of the dichroic band suggests the presence of a large amount of flexible organization for the peptide backbone. The CD spectrum has been analysed by computer program (Menendez-Arias *et al.*, 1988) in order to calculate the secondary structure content. An RMS value of 5 was calculated for the fit of the CD spectrum. The content of secondary structure is reported in Table II for comparison with prediction results as well as with previously published CD results. A global agreement appears between experimental and prediction results. In the Discussion, the differences will be evaluated in more detail.

Figure 6 reports the sulfydryl group titration of the protein by DTNB. Only three -SH groups out of four cysteine residues were detected in the native protein. All four groups were instead titrated in the presence of 5 M GdnHCl (not shown). These experiments suggest that one cysteine residue is deeply buried in the native protein structure and not available to the solvent. The two-phase titration curve of the native protein indicates a fast reactive sulfydryl group which is rapidly titrated while the other two cysteines are titratable only at a higher DTNB concentration, probably because one of the titrated residues is very reactive and/or more exposed. A peptide mapping analysis, performed to verify the cysteine labelling, indicates Cys72 as the main target of the labelling reaction (data not shown). Comparisons of the features of the model to these experimental data, as well as to previously published results, are reported in the Discussion.

## Discussion

It is a widely shared opinion that the sequence alignment represents the critical point of comparative modelling with low to medium sequence identity between the target and the template protein. In our work the alignment has been optimized by considering the alignments of sequences and secondary structure predictions of all homologous proteins found in the databases. Sequence alignments are based on parameters like the number of matches, scores for amino acid substitutions, gaps penalty, etc., whereas secondary structure predictions are based on different chemico-physical structural features of amino acids, depending on the different methods. We used the SOPMA predictions which consider the consensus of different independent methods. Therefore, we confirmed the good quality of the multiple alignment of sequences by the good correspondence of secondary structures, minimizing the possibility of

```
1          ----------------MADDLDFET-GDAGASATFPMQCSALRKNGFVVLKGRPCKIVEM-STSKTGKHGHAKVHLVGIDIFTGKKYEDI-CPSTHNMDVPNIKRNDFQLIGIQ-DGYLSLLQ-DSGEVREDLRLP-EGDLGK--EIEQKYDCGE-EILITVLSAMTEEAAVAIKAMA-K----
2          ----------------MADDLDFET-GDAGASATFPMQCSALRKNGFVVLKGRPCKIVEM-STSKTGKHGHAKVHLVGIDIFTGKKYEDI-CPSTHNMDVPNIKRNDFQLIGIQ-DGYLSLLQ-DSGEVREDLRLP-EGDLGK--EIEQKYDSGE-EILITVLSAMTEEAAVAIKAMAK-----
3          ----------------MADDLDFET-GDAGASATFPMQCSALRKNGFVVLKGWPCKIVEM-SASKTGKHGHAKVHLVGIDIFTGKKYEDI-CPSTHNMDVPNIRRNDFQLIGIQ-DGYLSLLQ-DSGEVPEDLRLP-EGDLGK--EIEQKYDCGE-EILITVLSAMTEEAAVAIKAMAK-----
4          ----------------MADELDFTT-GDAGASSTYPMQCSALRKNGFVVLKGRPCKIVEM-STSKTGKHGHAKVHLVGIDIFNGKKYEDI-CPSTHNMDVPNIKRNDYQLIGIQ-DGYLSLLT-ESGEVREDLKLP-EGDLGK--EIEGKFNANE-DVQISVISAMNEECAVAIKPCK------
5          --------------MADIEDTHFET-GDSGASATFPMQCSALRKNGFVMLKGRPCKIVEM-STSKTGKHGHAKVHLVGIDIFNGKKYEDI-CPSTHNMDVPHVKREDYQLTDIS-DDGYLTLMADNGDLREDLKIP-DGDLGT--QLRSDFDSGK-ELLCTVLKSCGEECVIAVKANTALD-K-
6          --------------MAELDD-QFET-TDSGASTTYPMQCSALRKNGFVMLKSRPCKIVEM-STSKTGKHGHAKVHMVGIDIFSNKKYEDI-CPSTHNMDVPNVKREDLQLIAIS-DDSFLTLMTESGDLREDLKVP-EGELGE--QLRLDFDSGK-DLLCTVLKACGEECVIAIKTNTALDK--
7          --------------MAEEEHVDFEG-GEAGASLTFPMQCSALRKNGHVVIKGRPCKIVDM-STSKTGKHGHAKVHIVALDIFNGRKYEDM-SPSTHNMDVPVVKRDEYQLVNID-DGYLNLMT-TDGTTKDDVRLP-EGELGN--EIEEGFDEGR-DLIITVVSAMGEETALACRDAPS-S---
8          --------------MAEEEHVDFEG-GEAGASLTFPMQCSALRKNGHVVIKGRPCKIVDM-STSKTGKHGHAKVHIVALDIFNGRKYEDM-SPSTHNMDVPVVKRDEYQLVNID-DGYLNLMT-TDGTTKDDVRLP-EGELGN--EIEEGFEEGK-DLIITVVSAMGEEIALACRDAPSA----
9          ----------------MAEEDTFET-ADAGAALTFPMQCSALRKNGHVVIKNRPCKIVDM-STSKTGKHGHAKVHLVAIDIFTGKKLEDL-SPSTHNMEVPNVSRQEFQLLDID-DGYLSLMT-ADGDTKDDVKVP-EGELGD--KLQSEFDEGK-DLIVTIISAMGEEAAISYKEAPKGSA--
10         ---------------MSDEEHTFET-ADAGSSATYPMQCSALRKNGFVVIKSRPCKIVDM-STSKTGKHGHAKVHLVAIDIFTGKKLEDL-SPSTHNMEVPVVKRNEYQLLDID-DGFLSLMN-MDGDTKDDVKAP-EGELGD--SLQTAFDEGK-DLMVTIISAMGEEAAISFKEAART--D-
11         ---------------MSDEEHTFEN-ADAGASATYPMQCSALRKNGFVVIKGRPCKIVDM-STSKTGKHGHAKVHLVTLDIFTGKKLEDL-SPSTHNLEVPFVKRSEYQLLDID-DGYLSLMT-MDGETKDDVKAP-EGELGD--SMQAAFDEGK-DLMVTIISAMGEEAAISFKEAPRSD---
12         -MKPLIMEYNKMSDNEALDVEDYAQ-AGSGASLTFPIQCSALRKNGFVVIKGFPCKIVDM-STSKTGKHGHAKVNITAIDIFTGKKYEEI-CPSTHNIDVPNVSRKEYTVMDVQ-DGYLSLLD-AGGEVKEDLALP-EDDIGK--EITQMLKEGK-EPLVSVISALGKEGVVSVKVSNN-----
13         -------------------------------------QCSALRKNGFVVLKGRPCKXXXX-STSKTGKHGHAKVHLVGIDIFTGKKYEDI-CPSTHNMDVPNIKRNDFQLIGIQ-DGFLSLLQ-DSGEVREDLRLP-EGELGR-----------------------------------------
14         ---------MSAAHDDAQHEHTFDS-VDAGRSATYPMQCSALRKGGHVVIKNRPCKIVDM-STSKTGKHGHAKVHLVAIDIFTGKKLEDL-CPSTHNMDVPNVKRTDYQFSYID-EDFLVLID-SNGEEKRELKMP-EGELAK--RIEKLEEEGK-DFFVGVQTAMGEEAAIDVKEASNKD---
15         ------------MSDDHHDDEHFHT-GDSGAAATFPKQCSALRKNEHVMIKGRPCKIVEM-STSKTGKHGHAKVHMVAIDIFTSKKLEDI-CPSTHNMDVPVVKRREYLLMAID-DGYCSLMDPESCEQKDDLKLP-DTELGQ--QIRDAYEKDEGSVLVQVVSAIGEEAILGWKVSTKE----
16         MARVAAPKTIESMSEDHHDEEQFDS-AESGAAATFPKQCSALRKNEHVMIRGRPCKIVEM-STSKTGKHGHAKVHMVAIDIFTTKKLEDI-CPSTHNMDVPVVKRREYILMSIE-DGFCSLMDPESCELKDDLKMP-EGDLGN--TIREALEKDEGSVLVQVVAACGEEAILGYKISTK----E
17         ------------MSEDHHDEEQFDS-AESGAAATFPKQCSALRKNEHVMIRGRPCKIVEM-STSKTGKHGHAKVHMVAIDIFTTKKLEDI-CPSTHNMDVPVVKRREYILMSIE-DGFCSLMDPESCELKDDLKMP-EGDLGN--TIREALEKDEGSVLVQVVAACGEEAILGYKISTKE----
18         -------------------------------------QXSALRKNGFVVLKGRPXK-----STSKTGKHGHAKVHLVGIDIFTGKKYEDI-XPSTHNMDVPNIKRNDFQLIGIQ-DGFLSLLQ-DSGEVREDLRLP-EGELG---R--------------------------------------
19         --------------MSD-EEHHFES-SDAGASKTYPQQAGTIRKNGYIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTSKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGYVSLLT-DNGSTKDDLKLPNDDTLLQ--QIKSGFDDGK-DLVVSVMSAMGEEQINALKDIGP---K-
20         --------------MSD-EEHHFESKADAGASKTYPQQAGTIRKNGYIVIKGRPCKVVEV-STSKTGKHGHAKCHFVAIDIFNGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ESGNTKDDLRLPTDESLLK--QVKDGFQEGK-DLVVSVMSAMGEEQINAIKDIGTK-N--
21         --------------MSD-EEHQFESKADAGASKTYPQQAGTIRKSGYIVIKGRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-DNGNTKDDLRLPTDDSLLS--QIKDGFAEGK-DLVVSVMSAMGEEQINALKDIGPK----
22         --------------MSD-EEHQFESKADAGASKTYPQQAGTIRKNGHIVIKGRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDDNLLT--QIKDGFAEGK-DLVVSVMSAMGEEQICALKDIGP-K---
23         ------------------HEDHFESKADAGASKTYPQQAGTIRKNGYIVIKNRPCKVVEV-STSKTGKHGHAKCHFVGIDIFNGKKLEDI-VPSSHNCDVPHVTRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDESLLS--QIKDGFGDGK-DLVVTVMSSMGEEQICALKDIGPK----
24         --------------MSDSEEHHFESKADAGASKTYPQQAGTIRKNGYIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFNGKKLEDI-VPSSHNCDVPHVNRTEYQLIDISEDGFVSLLT-ESGNTKDDLRLPTDDSLLG--QIKTGFGEGK-DLVVTVMSAMGEEQICALKDIGP-K---
25         --------------MSD-EEHHFESKADAGASKTYPQQAGTIRKSGYIVIKGRPCKVVEV-STSKTGKHGHAKCHFVAIDIFNGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDDALLN--QVKGGFEEGK-DLVLSVMSAMGEEQICAVKDIGTKS---
26         --------------MSDSEEHHFESKADAGASKTYPQQAGTVRKNGFIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFNGKKLEDI-VPSSHNCDIPHVNRTEYQLIDISEDGFVSLLT-SDGNTKDDLRLPTDETLVA--QIKEGFESGK-DLVVTVQSAMGEEQICALKDVGPK----
27         --------------MSD-EEHHFESKADAGASKTYPQQAGTIRKNGHIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDDNLLA--LIKDGFAEGK-DLVLSVMSAMGEEQICGIKDVGP-K---
28         --------------MSD-EEHHFESKADAGASKTYPQQAGTIRKSGHIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDDTLLA--QVKDGFAEGK-DLVLSVMSAMGEEQICGIKDIGPK----
29         -----------------------------GASKTYPQQAGTIRKNGHIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLRLPTDDNLLA--LIKDGFAEGK-DLVLSVMSAMGEEQICGIKDVGPK----
30         --------------MSD-EEHQFESKADAGASKTYPQQAGTIRKGGHIVIKNRACKVVEV-STSKTGKHGHAKCHFVAIDIFNGKKLEDI-VPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLKLPTDDALLT--QIKDGFGEGK-DLVVSVMSAMGEEQICALKDIGPK----
31         --------------MSD-EEHQFESKADAGASKTYPQQAGTIRKNGYIVIKNRPCKVVEV-STSKTGKHGHAKCHFVAIDIFTSKKLEEVYVPSSHNCDVPHVNRTDYQLIDISEDGFVSLLT-ENGNTKDDLKLPTDDSLLT--QIKDGFAEGK-DLVVSVMSAMGEEQICALKDIGGKN---
32         -------------------------------------------------------------STSKTGXHGHAKVHLVGIDIFTGKKYEDI-XPSTHNMDVPNIKRNDFQLIGIQ-DGFLSLLQ-DSGEVREDLRLP-EGELGR-----------------------------------------
33         ----------------------FTT-GDAGASSTYPMQXSALRKNGFVVLKGRPXKIVEM-STSKTGXHGHAKVHLVGIDIFNGKKYEDI-XPSTHNMDVPNIK--------------------------------------------------------------------------------
34         ------------------------M-GDFGASKTLPKQCSALRKGGHVVMKDRPCKIVEM-STSKTGKHGSAKVHLVG----------------------------------------------------------------------------------------------------------
35         ---------------------------MIIMPGTKQVNVGSLKVGQYVMIDGVPCEIVDI-SVSKPGKHGGAKARVVGIGIFEKVKKEFV-APTSSKVEVPIIDRRKGQVLAIM-GDMVQIMD---LQTYETLELPIP-------EGIEGLEPGG---EVEYIEAVGQYKITRVIGGK------
36         ------------------------------MPGTKQVNVGSLKVGQYVMIDGVPCEIVDI-SVSKPGKHGGAKARVVGIGIFEKVKKEFV-APTSSKVEVPIIDRRKGQVLAIM-GDMVQIMD---LQTYETLELPIP-------EGIEGLEPGG---EVEYIEAVGQYKITRVIGGK------
37         --------------------------MVIIMPGTKQVNVGSLKVGQYVMIDGVPCEIVDI-SVSKPGKHGGAKARVVGIGIFEKVKKEFV-APTSSKVEVPIIDRRKGQVLAIM-GDMVQIMD---LQTYETLELPIP-------EGIEGLEPGG---EVEYIEAVGQYKITRVIGG---K---
38         ---------------------------VIIMPGTKQVNVGSLKVGQYVMIDGVPCEIVDI-SVSKPGKHGGAKARVVGIGIFEKVKKEFV-APTSSKVEVPIIDRRKGQVLAIM-GDMVQIMD---LQTYETLELPIP-------EGIEGLEPGG---EVEYIEAVGQYKITRVIGGK------
39         ---------------------------------------------------------------------------------------------------------NDYQLIGIQ-DGYLSLLT-ESGEVREDLKLP-EGDLGK-----------------------------------------
40         ----------------------MPSSFGCSRMSVNYATLGDLKKGSYIVIDGEPCRIVEM-SRAKTGKHGSAKAHVVAICVFSGQKKSLV-APVDTRVQIPVIEKRLGQVLADM-GDMVQIMD---LETYDTFEVEKPGGNEEEEQLAAKLQPGV---TVEYWLIMGKPKIIRIRSSSS-----
41         ---------------------------MGIQMSIQYTTVGDLKVGSYVMIDGEPCRVVEI-TKAKTGKHGSAKANVVAIGLFTGQKRSLM-APVDQQVEVPIIEKHVGQILADK-GDNLTIMD---LESYETFDLEKP--TEN--EIVSKIRPGA---EIEYWSVMGRRKIVRVK---------
42         --------------------------------MKQQVEVRQLREGGYVVIDDEPCEILSI-SVSKPGKHGAAKARIDAIGIFDSQKRSIV-QPVTAKIYVPIVERKRAQIISVT-GNVAQIMD---LETYETFELEVP----E--ELKDKMEQGR---EVIYLESLGKRKIERMA---------
43         ----------------------------KWVMSTKYVEAGELKEGSYVVIDGEPCRVVEI-EKSKTGKHGSAKARIVAVGVFDGGKRTLS-LPVDAQVEVPIIEKFTAQILSVS-GDVIQLMD---MRDYKTIEVPMKYVEEE---AKGRLAPGA---EVEVWQILDRYKIIRVK-G-------
44         --MGYYDDDAHGHVEADAAPRATTGTGTGSASQIVTIPCHHIRLGDILILQGRPCQVIRISTSAATGQHRYLGVDLFTKQLHEESSFVSNPAPSVVVQTMLGPVFKQYRVLDMQ-DGSIVAMT-ETGDVKQNLPVIDQSSLWN--RLQKAFESGRGSVRVLVVSDHGREMAVDMKVVHGSRL--
45         ----------------------------KWVXSTKYVEAGELKEGSYVVIDGEPCRVVEI-EKSKTGKHGSAKARIVAVGVFDGGKRTLS-LPVDAQVEVPIIEKFTAQILSVS-GDVIQLXD---XRDYKTIEVPXKYVEEE---AKGRLAPGA---EVEVWQILDRYKIIRVKG--------
46         ------------------------------MGDKTKVQVSKLKPGRYIIIDDEPCRIVNI-TVSSPGKHGSAKARIEAVGIFDGKVRSIV-KPTSAEVDVPIIDKKTAQVIAIT-PDTVQIMD---METYEMFEVPIDTGVAE--EIRDQLKEGI---NVEYWETLGRVKIMRIKGEG-E----
47         ------------------------------MGDKTKVQVSKLKPGRYIIIDDEPCRIVNI-TVSSPGKHGSAKARIEAVGIFDGKVRSIV-KPTSAEVDVPIIDKKTAQVIAIT-PDTVQIMD---METYETFEVPIDTGVAD--EIRDQLKEGI---NVEYWETLGRIKIMRIKGEGE-----
48         -------------------------------MSKKVVEVKTLKVGKYVIIDGEASKITNI-STSSPGKHGSAKARVEAVGIFDNQKRSFV-KPVDSKVDIPIIDKRTAQVIAIM-GGDVQLMD---LETYETFETPIP----D--ELSEQLVEGV---EVEYIEALGQRKLMRTKG--------

Consensus:                              GA  T P Q   LRK G VVIKGRPCK V  -STSKTGKHGHAK H VAIDIF GKK ED - PS HN DVP   R   QL  I -DG    L -  G    DL LP    L  --        G -   V V  A G E     K
```

**Fig. 2.** Multiple alignment of 48 eIF-5A sequences. The alignment was created by the CLUSTALW program.

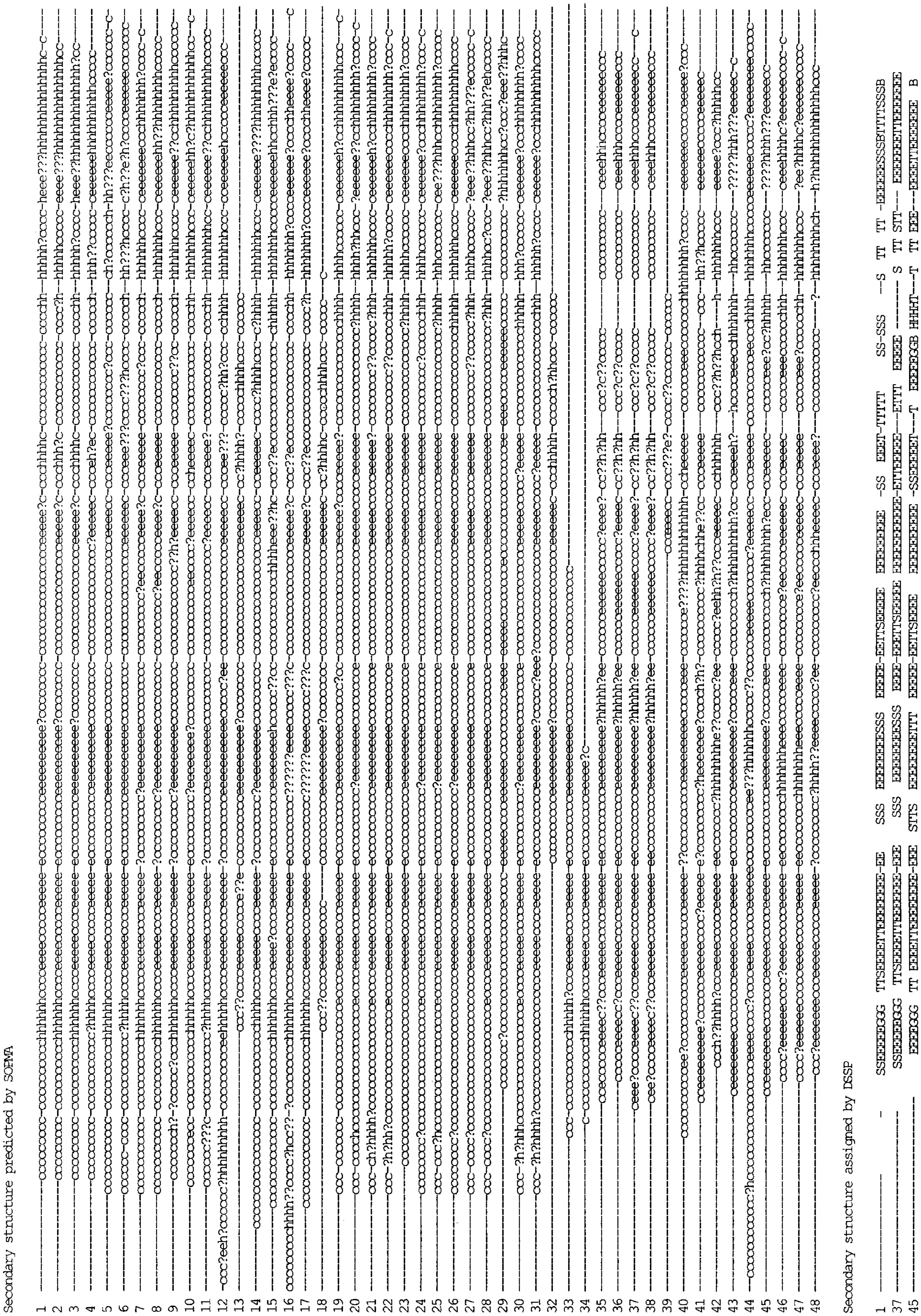

Secondary structure predicted by SOPMA
Secondary structure assigned by DSSP

**Fig. 3.** Alignment of the predicted secondary structure of the 48 proteins and assigned by DSSP to the experimental models of the two archea proteins and to the predicted model of the human protein. Letter code: lower-case refers to prediction by SOPMA, upper-case refers to DSSP assignment for the predicted 3D model of the human eIF-5A and the experimental models for archea proteins. H/h, helix; E/e, beta strand; c, coil; ?, ambiguous state; T, turn; G, 3/10 helix; S, bend; B, beta bridge.

**Table II.** Secondary structure content

| | Alpha helix (%) | Beta structure (%) | Turn[a] (%) | Other[a] (%) |
|---|---|---|---|---|
| Experimental results | | | | |
| CD of human eIF-5A (this paper) | 8 ± 3 | 21 ± 3 | 20 ± 2 | 50 ± 3 |
| CD of human eIF-5A (Klier *et al.*, 1995) | 10 | 38 | Not reported | 52 |
| 3D experimental model of *M.jannaschii* IF5A | 0 | 58 | 13 | 29 |
| 3D experimental model of *P.aerophilum* IF5A | 3 | 50 | 15 | 31 |
| Predictions | | | | |
| 3D model of human eIF-5A (structure assigned by DSSP) | 0 | 35 | 13 | 52 |
| SOPMA prediction for human eIF-5A | 20 | 19 | Not reported | 61 |
| PHD prediction for human eIF-5A | 6 | 35 | Not reported | 49 |
| JPRED prediction for human eIF-5A | 11 | 36 | Not reported | 43 |

[a]When the 'turn' structure is not reported, it must be considered as enclosed in the 'other' structure.

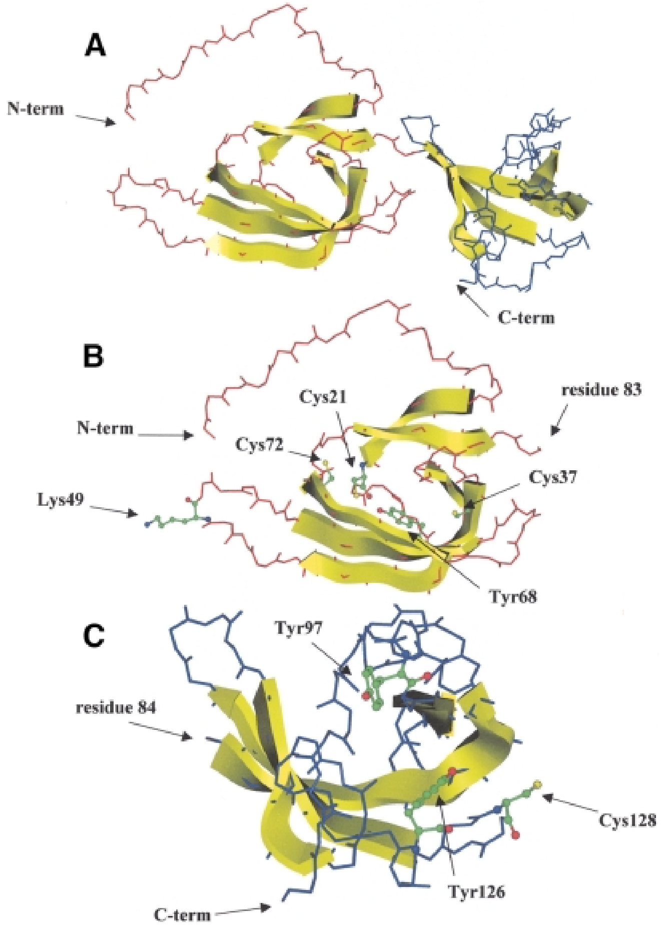


**Fig. 4.** 3D model of human eIF-5A. The schematic view of the backbone evidences the presence of two distinct domains (N-terminal: red; C-terminal: blue) consisting of mainly beta structure (yellow arrows). The N-terminal domain is a 'roll' architecture. Lys49, cysteine and tyrosine amino acids have been highlighted. Residue 83 is the last of this domain. The C-terminal domain is folded in a 'barrel' architecture. Cysteine and tyrosine amino acids have been highlighted. Residue 84 is the first of this domain.

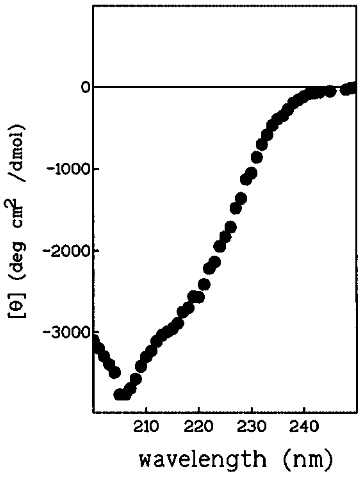


**Fig. 5.** Far-UV CD spectra of eIF-5A. See Methods for experimental details.

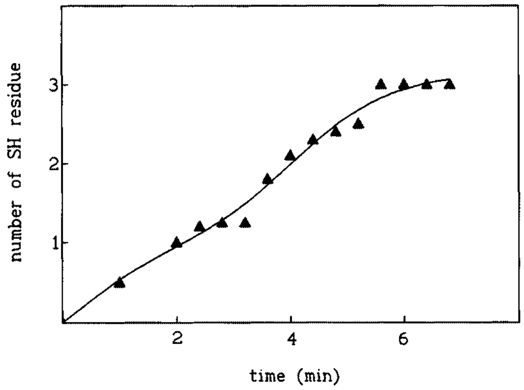


**Fig. 6.** Titration of cysteinyl residues. See Methods for experimental details.

errors in the homology modelling procedure. We also verified that the gaps do not break secondary structure elements observed in the experimental models nor in the predictions, in order to prevent a mistake in sequence alignment. Ten predicted models were checked for their stereo-chemical quality and the best one was selected as our final model. The submission of the model to the PDB, and its acceptance, has provided further validation of the good quality of our model. Last, we performed experimental measurements of spectral features, in order to compare the structural feature of the final model to our experimental evidence. In fact, in our opinion the structural prediction should be always verified on the basis of experimental results, particularly when the degree of sequence similarity is under the 40% threshold.

The secondary structure content of the human eIF-5A, detected by far-UV CD, suggests that the protein can be classified as an all-beta protein, as confirmed by the crystallographic models from archea and their classification in the

CATH (Orengo *et al*., 1997) and SCOP databases (Murzin *et al*., 1995). As is evident from data reported in Table II, the comparison of experimental data show that differences of secondary structure contents exist between human eIF-5A (CD spectra) and the archea homologues (crystallographic data). The most evident differences are the lower content of beta structure and the higher content of unordered structure (other) in the human protein. The same trend appears between the predicted 3D model of the human protein and the 3D experimental models of archea proteins. This means that the archea structures, used as templates for the homology modelling, have been correctly modified in the modelling procedure, and the final model is in good agreement with the expected differences between the human and the archea proteins. The few differences in the percentage content of secondary structure between CD data and the model can be ascribed, at least in part, to the presence of an unordered segment in the N-terminal region of the human sequence, without correspondence in the template proteins, and therefore not folded by the homology modelling. Secondary structure predictions assign coil conformation to this segment, but it is possible that it partially assumes helical conformation, thus explaining the presence of an alpha helix detected by the CD spectrum, although to a very low extent. The secondary structure predictions, on the contrary, show higher differences to the experimental results, thus confirming that the homology modelling approach, applied in suitable conditions, represents the best method for protein structure prediction.

Comparisons of detailed structural features of the modelled protein to experimental results also confirms the reliability of the model. Cysteine titration by DTNB and labelling by fluorophore constitute a complex set of information. In fact, experimental results indicate that: (i) all four cysteines are titrated when the protein is in the presence of a high denaturant concentration; (ii) only one cysteine is quickly titrated when the protein is in native conditions; (iii) the other two cysteines are titrated after a longer incubation time; and (iv) the fourth cysteine is not titratable. The first cysteine titrated should be Cys72, as obtained by tryptic mapping data. Moreover, as previously published (Stiuso *et al*., 1999), the *N*-(iodoacetylaminoethyl)-5-naphtylamine-1-sulfonic acid (IAEDANS) labelling of cysteine side chains, followed by tryptic mapping, revealed that Cys72 is evidently marked and the spectral features of the labelled protein suggest that the IAEDANS bounded is in an apolar environment. These results can be explained by accurate evaluation of the cysteine environment in the predicted model. In Figure 4 the cysteine residues have been highlighted. Cys128 appears well exposed on the surface of the protein and the analysis of the model with specific software confirms that 69% of the amino acid surface is accessible to the solvent, whereas Cys72 has an intermediate exposure (9%). Cys21 and Cys37 are buried, their solvent accessibility being 0.8 and 0.4%, respectively. It could be expected that it can be much easier to titrate and label the most exposed Cys128 than Cys72. However, the apolar environment detected for the IAEDANS fluorophore seems not to be suitable for the high exposure of Cys128 because the fluorophore, after the labelling of such a side chain, might be very exposed to the solvent. Cys72 is partially buried and this can justify the apolar environment of the IAEDANS. The preference for Cys72 instead of Cys128 may be the consequence of other environmental conditions, like the presence of specific functional groups which facilitate or prevent the labelling.

It was previously shown (Stiuso *et al*., 1999) that human eIF-5A consists of two different domains which seem to react differently to the presence of denaturant. The fluorescence of tyrosine side chains, mainly present in the C-terminal domain (Figure 4C), show a structural transition at low denaturant concentration (mid-point at approximately 1 M GdnHCl) whereas the IAEDANS-labelled protein, i.e. the protein with derivatized cysteines, mainly present in the N-terminal domain (Figure 4B) seems to be more stable, its fluorescence being subjected to a transition at a higher GdnHCl concentration (mid-point at 2 M). This information suggested that the N-terminal domain is more stable than the C-terminal domain. This statement is in good agreement with our structural model, in which two distinct domains are evident and the N-terminal domain has a lower amount of unordered structure, which can mean a more compact and rigid backbone structure.

In the past, secondary structure predictions allowed one to hypothesize the nature of the packing of the two domains (Klier *et al*., 1995). However, such a hypothesis appears to be strongly affected by the excessive alpha helix prediction, evidently overestimated as compared to both our CD spectra analysis and CD published results of the same authors (Joao *et al*., 1995; Klier *et al*., 1995). In a more recent predictive paper (Gerloff *et al*., 1998) a lower content of helix was predicted by using the newest available methods. In the absence of homologue structures, the fold recognition approach was applied and mainly beta folds were suggested, like the open barrel and the immunoglobulin-like beta sandwich. Finally, as a further improvement, our prediction gives a 3D model of the human eIF-5A, based on the comparative modelling approach applied after a careful adjustment of the sequence alignment. This structural model of eIF-5A can be useful for designing ligands able to inhibit the lysine→hypusine modification, thus acting as new antiproliferative agents raised against eIF-5A in order to potentiate the apoptosis induced by IFNα in cancer cells. In fact, we have demonstrated that the addition of IFNα to human epidermoid cancer KB cells induces apoptosis (Caraglia *et al*., 1999) and reduces hypusine synthesis (Caraglia *et al*., 1995, 1999). Such effects are both antagonized when IFNα-treated KB cells are exposed to EGF for 12 h (Caraglia *et al*., 2000). Therefore, on the basis of these results, hypusine synthesis could be part of an anti-apoptotic response raised by EGF IFNα-treated cells. The knowledge of the eIF-5A structure could allow a pharmacological screening on a computational basis in order to identify pharmacological substances able to bind the hypusine containing site of eIF-5A and inhibit its function. Therefore, eIF-5A, on the basis of its intrinsic biochemical and structural properties, could represent an useful target in combined approaches for the inhibition of tumour cell proliferation (Caraglia *et al*., 2000).

In conclusion, the comparison of experimental data and structural features highlighted on the predicted model allows us to validate both the homology modelling strategy, based on an accurate refinement of the sequence alignment, and the predicted model. This will be useful for further investigations in pharmacological studies aimed at modulating the eIF-5A function.

## Acknowledgements

This work has been partially funded by grants from MURST, CNR and Regione Campania to A.A. and G.C.

## References

Abbruzzese,A. (1988) *J. Neurochem.*, **50**, 695–699.

Abbruzzese,A., Park,M.H. and Folk,J.E. (1985) *Fed. Proc.*, **44**, 1487.

Abbruzzese,A., Park,M.H. and Folk,J.E. (1986) *J. Biol. Chem.*, **261**, 3085–3089.

Abbruzzese,A., Liguori,V. and Park,M.H. (1988a) *Adv. Exp. Med. Biol.*, **250**, 459–466.

Abbruzzese,A., Isernia,T., Liguori,V. and Beninati,S. (1988b) In Perin,A., Scalabrino,G., Sessa,A. and Ferrioli,M.E. (Eds), *Perspectives in Polyamine Research.* Wichtig, Milan, Italy, pp. 79–84.

Abbruzzese,A., Park,M.H., Beninati,S. and Folk,J.E. (1989) *Biochem. Biophys. Acta*, **997**, 248–255.

Abbruzzese,A., Hanauske-Abel,H.M., Park,M.H., Henke,S. and Folk,J.E. (1991) *Biochem. Biophys. Acta*, **1077**, 159–166.

Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.

Beninati,S., Abbruzzese,A. and Folk,J.E. (1990) *Ann. Biochem.*, **184**, 16–20.

Beninati,S., Abbruzzese,A. and Cardinale,M. (1993) *Int. J. Cancer*, **53**, 792–797.

Beninati,S., Nicolini,L., Jakus,J., Passeggio,A. and Abbruzzese,A. (1995) *Biochem. J.*, **305**, 725–728.

Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.

Caraglia,M., Passeggio,A., Beninati,S., Leardi,A., Nicolini,L., Improta,S., Pinto,A., Bianco,A.R., Tagliaferri,P. and Abbruzzese,A. (1997) *Biochem. J.*, **324**, 737–741.

Caraglia,M. *et al.* (1999) *Cell Death Differ.*, **6**, 773–780.

Caraglia,M, Budillon,A., Vitale,G., Lupoli,G., Tagliaferri,P. and Abbruzzese A. (2000) *Eur. J. Biochem.*, **267**, 3919–3936.

Chung,S.I., Park,M.H., Folk,J.E. and Lewis,M. (1991) *Bioch. Biophys Acta*, **1076**, 448–451.

Cooper,H.L., Park,M.H. and Folk,J.E. (1982) *Cell*, **29**, 791–797.

Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M. and Barton,G.J. (1998) *Bioinformatics*, **14**, 892–893.

Ellman,G.L. (1959) *Arch. Biochem. Biophys*, **82**, 70–77.

Flores,T.P., Orengo,C.A., Moss,D.S. and Thorton,J.M. (1993) *Protein Sci.*, **2**, 1811–1826.

Geourjon,C. and Deleage,G. (1994) *Protein Eng.*, **7**, 157–164.

Geourjon,C. and Deleage,G. (1995) *Comput. Appl. Biosci.*, **11**, 681–684.

Gerloff,D.L., Joachimiak,M., Cohen,F.E., Cannarozzi,G.M., Chamberlin,S.G. and Benner,S.A. (1998) *Biochem. Biophys. Res. Commun.*, **251**, 173–181.

Gill,S.C. and von Hippel,P.H. (1989) *Anal. Biochem.*, **182**, 319–326.

Gordon,E.D., Mora,R., Meredith,S.C., Lee,C. and Lindquist,S.L. (1987) *J. Biol. Chem.*, **262**, 16585–16589.

Hershey,J.W.B. (1991) *Annu. Rev. Biochem.*, **60**, 717–755.

Hershey,J.W.B., Smit-McBride,Z. and Schnier,J. (1990) *Biochim. Biophys Acta*, **1050**, 160–162.

Hubbard,S.J., Campbell,S.F. and Thornton,J.M. (1991) *J. Mol. Biol.*, **220**, 507–630.

Jakus,J., Wolff,E.C., Park,M.H. and Folk,J.E. (1993) *J. Biol. Chem.*, **268**, 13151–13159.

Joao,H.C., Csonga,R., Klier,H., Koettnitz,K., Auer,M. and Eder,J. (1995) *Biochemistry*, **34**, 14703–14711.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kim,K.K., Hung,L.W., Yokota,H., Kim,R. and Kim,S.H. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 10419–10424.

Klier,H., Csonga,R., Joao,H.C., Eckerskorn,C., Auer,M., Lottspeich,F. and Eder,J. (1995) *Biochemistry*, **34**, 14693–14702.

Lalande,M. and Hanausske-Abel,H.M. (1990) *Exp. Cell. Res.*, **188**, 117–121.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Martin,A.C., MacArthur,M.W. and Thornton,J.M. (1997) *Proteins*, Suppl. 1, 14–28.

Menendez-Arias,L., Gomez-Gutierrez,J., Garcia-Fernandez,M., Garcia-Tejedor,.A. and Moran,F. (1988) *Comput. Applic. Biosci.*, **4**, 479–482.

Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) *J. Mol. Biol.*, **247**, 536–540.

Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.M. (1997) *Structure*, **5**, 1093–1108.

Park,M.H. (1987) *J. Biol. Chem.*, **262**, 12730–12734.

Park,M.H., Cooper,H.L. and Folk,J.E. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 2869–2873.

Park,M.H., Cooper,H.L. and Folk,J.E. (1982) *J. Biol. Chem.*, **257**, 7219–7222.

Park,M.H., Chung,S.I., Cooper,H.L. and Folk,J.E. (1984) *J. Biol. Chem.*, **259**, 4563–4565.

Park,M.H., Wolff,E.C., Smith-McBride,Z., Hershey,J.W.B. and Folk,J.E. (1991) *J. Biol. Chem.*, **266**, 7988–7994.

Park,M.H., Wolff,E.C. and Folk,J.E. (1993) *Trends Biochem. Sci.*, **18**, 475–479.

Peat,T.S., Newman,J., Waldo,G.S., Berendzen,J. and Terwilliger,T.C. (1998) *Structure*, **6**, 1207–1214.

Rodriguez,R., Chinea,G., Lopez,N., Pons,T. and Vriend,G. (1998) *Bioinformatics*, **14**, 523–528.

Rosorius,O., Reichart,B., Kratzer,F., Heger,P., Dabauvalle,M.C. and Hauber,J. (1999) *J. Cell. Sci.*, **112**, 2369–2380.

Rost,B. and Sander,C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7558–7562.

Sali,A. and Blundell,.L. (1993) *J. Mol. Biol.*, **234**, 779–815.

Schnier,J., Schwelbeerger,H.G., Smit-McBride,Z., Kang,H.A. and Hershey,J.W.B. (1991) *Mol. Cell. Biol.*, **11**, 3105–3114.

Smit-McBride,Z., Dever,T.E., Hershey,J.W.B. and Merrick,W.C. (1989) *J. Biol. Chem.*, **264**, 1578–1583.

Sternberg,M.J., Bates,P.A., Kelly,L.A. and MacCallum,R.M. (1999) *Curr. Opin. Struct. Biol.*, **9**, 368–373.

Stiuso,P., Colonna,G., Ragone,R., Caraglia,M., Hershey,J.W., Beninati,S. and Abbruzzese,A. (1999) *Amino Acids*, **16**, 91–106.

Thompson,J.D., Higgins,D.G., Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.

Tome,M.E. and Gerner,E.W. (1997) *Biol. Signals*, **6**, 150–156.

Tome,M.E., Fiser,S.M., Payne,C.M. and Gerner,E.W. (1997) *Biochem. J.*, **328**, 847–854.

Westhead,D.R. and Thornton,J.M. (1998) *Curr. Opin. Biotech.*, **9**, 383–389.

Wetlaufer,D.B. (1962) *Adv. Protein Chem.*, **17**, 303–390.

*Received February 19, 2001; revised July 27, 2001; accepted August 9, 2001*